JOEL BEST

STAT-SPOTTING

A FIELD GUIDE TO IDENTIFYING DUBIOUS DATA

"An ideal guide for anyone who reads a newspaper, watches television, or surfs the Web. In short, everyone." —JOE SWINGLE, *NUMERACY*

UPDATED AND EXPANDED

STAT-SPOTTING

STAT-SPOTTING

A FIELD GUIDE TO IDENTIFYING DUBIOUS DATA

UPDATED AND EXPANDED

JOEL BEST

University of California Press Berkeley Los Angeles London

University of California Press, one of the most distinguished university presses i the United States, enriches lives around the world by advancing scholarship in th humanities, social sciences, and natural sciences. Its activities are supported b the UC Press Foundation and by philanthropic contributions from individuals an institutions. For more information, visit www.ucpress.edu.

University of California Press
Berkeley and Los Angeles, California

University of California Press, Ltd.
London, England

© 2008, 2013 by The Regents of the University of California

First paperback printing 2013
ISBN 978-0-520-27998-8
eISBN 9780520957077

“Crystal-Clear Risk” on p. 23 is from “Party, Play—and Pay: Multiple Partners Unprotected Sex, and Crystal Meth,” *Newsweek,* February 28, 2005, 39. © 200 Newsweek, Inc. All rights reserved. Used by permission and protected by th Copyright Laws of the United States. The printing, copying, redistribution, c retransmission of the Material without express written permission is prohibited.

Earlier discussions of some of the examples used in this book appeared in “Birds- Dead and Deadly: Why Numeracy Needs to Address Social Construction *Numeracy* 1, no. 1 (2008): article 6, and in the “Out of Context” feature in issue of *Contexts* magazine published between spring 2005 and fall 2007 (volumes 4 6).

The Library of Congress has catalogued an earlier edition as follows:

Library of Congress Cataloging-in-Publication Data

Best, Joel.
Stat-spotting : a field guide to identifying dubious data / Joel Best.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-520-25746-7 (cloth : alk. paper)
1. Sociology—Statistical methods. 2. Social problems—Statistical methods. Title.
HM535.B477 2008
301.072’7—dc22 2008017175

Designer: Nola Burger

Text: 11/15 Granjon
Display: Akzidenz Grotesk and Akzidenz Grotesk Condensed

Manufactured in the United States of America

21 20 19 18 17 16 15 14 13
10 9 8 7 6 5 4 3 2 1

The paper used in this publication meets the minimum requirements of ANSI/NIS
Z39.48-1992 (R 2002) (*Permanence of Paper*).

The statistics you don’t compile never lie.

– STEPHEN COLBERT

CONTENTS

PREFACE TO THE 2013 EDITION

PART 1 GETTING STARTED

A. SPOTTING QUESTIONABLE NUMBERS

B. BACKGROUND

- B.1 Statistical Benchmarks
- B.2 Severity and Frequency

PART 2 VARIETIES OF DUBIOUS DATA

C. BLUNDERS

- C.1 The Slippery Decimal Point
- C.2 Botched Translations
- C.3 Misleading Graphs
- C.4 Careless Calculations

D. SOURCES: WHO COUNTED–AND WHY?

- D.1 Big Round Numbers
- D.2 Hyperbole
- D.3 Shocking Claims
- D.4 Naming the Problem

E. DEFINITIONS: WHAT DID THEY COUNT?

- E.1 Broad Definitions
- E.2 Expanding Definitions
- E.3 Changing Definitions
- E.4 The Uncounted

F. MEASUREMENTS: HOW DID THEY COUNT?

- F.1 Creating Measures

F.2 Odd Units of Analysis

F.3 Loaded Questions

F.4 Raising the Bar

F.5 Technical Measures

G. PACKAGING: WHAT ARE THEY TELLING US?

G.1 Impressive Formats

G.2 Misleading Samples

G.3 Convenient Time Frames

G.4 Peculiar Percentages

G.5 Selective Comparisons

G.6 Statistical Milestones

G.7 Averages

G.8 Epidemics

G.9 Correlations

G.10 Discoveries

H. RHETORIC: WHAT DO THEY WANT US TO THINK?

H.1 Using Short-Term Turnover to Measure Long-Term Problems

H.2 Sudden Turns for the Worse

H.3 Designating Myths

H.4 Rhetorical Flourishes

I. DEBATES: WHAT IF THEY DISAGREE?

I.1 Causality Debates

I.2 Equality Debates

I.3 Policy Debates

PART 3 STAT-SPOTTING ON YOUR OWN

J. SUMMARY: COMMON SIGNS OF DUBIOUS DATA

K. BETTER DATA: SOME CHARACTERISTICS

L. AFTERWORD: IF YOU HAD NO IDEA THINGS WERE THAT BAD, THEY PROBABLY AREN’T

M. SUGGESTIONS FOR THOSE WHO WANT TO CONTINUE STAT-SPOTTING

ACKNOWLEDGMENTS

INDEX

PREFACE TO THE 2013 EDITION

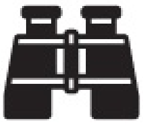
I was introduced to thinking about bad statistics when I read Darrell Huff's *Ho* *to Lie with Statistics* as a first-year college student. By the time I checked out th library's copy, that book was more than ten years old, and the volume showed lot of wear. Huff's examples no longer seemed current, and I particularl remember that the graphs he reproduced from newspaper articles struck me a somehow old-fashioned. Nonetheless, that book made a bigger impression on m than anything else I read during my first year in college. I understood that, even the book's examples were no longer timely, Huff's lessons were timeless. As th years went by, I found myself recalling *How to Lie with Statistics* when encountered dubious numbers in my reading. I discovered that people continue to get confused by statistics in the same ways Huff had identified.

Updating *Stat-Spotting* has forced me to think about what's timely and what timeless. This book offers a catalog of common problems that can be found i statistics, particularly the sorts of numbers that pop up when people are debatin social issues. These problems—the errors in reasoning people make when the present statistics—don't go away; people keep making the same mistake sometimes because they don't know any better, sometimes because they'r hoping to mislead an audience that isn't able to spot what's wrong with th numbers. These timeless lessons—knowing how to recognize some common way statistics can be flawed—are what I hope you will take from this book.

The book also contains examples meant to illustrate each of the errors describe. I chose most of these examples because they seemed timely (and, hoped, engaging). Inevitably, as the years pass, these examples age; they n longer seem to have been "ripped from the headlines." So, given the opportunit to update *Stat-Spotting,* I had to decide what needed to be replaced.

I chose to make some selective changes: (1) I revised all the benchmar statistics because a benchmark that isn't up-to-date isn't that useful; (2) I update the list of resources at the end of the book to give readers a better sense of wher they could find current information; (3) I added an entire section on the rhetorica uses of statistics, complete with new problems to be spotted and new example illustrating those problems; and (4) in a few cases where I knew the debate ha evolved in important ways, I revised the discussion of an example or added newe better references. However, most of the examples remain unchanged. They don strike me as antiquated, and I continue to feel that a lot of them are prett

interesting. More importantly, I hope that the people who read this book will focu
more on the timeless lessons and less on the timeliness of the examples.



PART 1

GETTING STARTED

A

SPOTTING QUESTIONABLE NUMBERS

The billion is the new million. A million used to be a lot. Nineteenth-centur Americans borrowed the French term *millionnaire* to denote those whose wealt had reached the astonishing total of a million dollars. In 1850, there were 2 million Americans; in the 1880 census, New York (in those days that mean Manhattan; Brooklyn was a separate entity) became the first U.S. city with mor than one million residents.

At the beginning of the twenty-first century, a million no longer seems all tha big. There are now millions of millionaires (according to one recent estimat about 8.6 million U.S. households have a net worth of more than $1 million, no counting the value of their principal residences).[1] Many houses are priced a more than a million dollars. The richest of the rich are billionaires, and even the are no longer all that rare. In fact, being worth a billion dollars is no longe enough to place someone on *Forbes* magazine's list of the four hundred riches Americans; some individuals have annual *incomes* exceeding a billion dollars Discussions of the U.S. economy, the federal budget, or the national debt speak o trillions of dollars (a trillion, remember, is a million millions).

The mind boggles. We may be able to wrap our heads around a million, bu billions and trillions are almost unimaginably big numbers. Faced with suc daunting figures, we tend to give up, to start thinking that all big numbers (sa everything above 100,000) are more or less equal. That is, they're all *a lot.*

Envisioning all big numbers as equal makes it both easier and harder to follo the news. Easier, because we have an easy way to make sense of the numbers Thus, we mentally translate statements like "Authorities estimate that HIV/AID kills nearly three million people worldwide each year" and "Estimates are that on billion birds die each year from flying into windows" to mean that there are *a lo* of HIV deaths and *a lot* of birds killed in window collisions.

But translating all big numbers into *a lot* makes it much harder to thin seriously about them. And that's just one of the ways people can be confused b statistics—a confusion we can't afford. We live in a big, complicated world, and w need numbers to help us make sense of it. Are our schools failing? What shoul we do about climate change? Thinking about such issues demands that we mov beyond our personal experiences or impressions. We need quantitative data-statistics—to guide us. But not all statistics are equally sound. Some of th numbers we encounter are pretty accurate, but others aren't much more tha wild guesses. It would be nice to be able to tell the difference.

This book may help. My earlier books—*Damned Lies and Statistics* and *Mor Damned Lies and Statistics*—offered an approach to thinking critically about th statistics we encounter.[3] Those books argued that we need to ask how number are socially constructed. That is, who are the people whose calculations produce the figures? What did they count? How did they go about counting? Why did the go to the trouble? In a sense, those books were more theoretical; they sought t understand the social processes by which statistics are created and brought t our attention. In contrast, this volume is designed to be more practical—it is field guide for spotting dubious data. Just as traditional field guides offer advic on identifying birds or plants, this book presents guidelines for recognizin questionable statistics, what I'll call "stat-spotting." It lists common problem found in the sorts of numbers that appear in news stories and illustrates eac problem with an example. Many of these errors are mentioned in the earlie books, but this guide tries to organize them around a set of practical question that you might ask when encountering a new statistic and considering whether might be flawed. In addition, all of the examples used to illustrate the variou problems are new; none appear in my other books.

This book is guided by the assumption that we are exposed to many statistic that have serious flaws. This is important, because most of us have a tendency t equate numbers with facts, to presume that statistical information is probab pretty accurate information. If that's wrong—if lots of the figures that w encounter are in fact flawed—then we need ways of assessing the data we'r given. We need to understand the reasons why unreliable statistics find their wa into the media, what specific sorts of problems are likely to bedevil thos numbers, and how to decide whether a particular figure is accurate. This book not a general discussion of thinking critically about numbers; rather, it focuses o common flaws in the sorts of figures we find in news stories.

I am a sociologist, so most of the examples I have chosen concern claims abou social problems, just as a field guide written by an economist might highligh dubious economic figures. But the problems and principles discussed in this boo are applicable to all types of statistics.

This book is divided into major sections, each focusing on a broad question such as: Who did the counting? or What did they count? Within each section, identify several problems—statistical flaws related to that specific issue. Th discussion of each problem lists some things you can "look for" (that is, warnin signs that particular numbers may have the flaw being discussed), as well as a example of a questionable statistic that illustrates the flaw. (Some of the example could be used to illustrate more than one flaw, and in some cases I note a example's relevance to points discussed elsewhere in the book.) I hope tha reading the various sections will give you some tools for thinking more critical about the statistics you hear from the media, activists, politicians, and othe advocates. However, before we start to examine the various reasons to suspec that data may be dubious, it will help to identify some statistical benchmarks tha can be used to place other figures in context.

B

BACKGROUND

Having a small store of factual knowledge prepares us to think critically abou statistics. Just a little bit of knowledge—a few basic numbers and one importa rule of thumb—offers a framework, enough basic information to let us begin t spot questionable figures.

B1 Statistical Benchmarks

When interpreting social statistics, it helps to have a rough sense of scale. Just few benchmark numbers can give us a mental context for assessing other figure we encounter. For example, when thinking about American society, it helps t know that:

- The U.S. population is something over 300 million (about 312 million i 2011).
- Each year, about 4 million babies are born in the United States (the 201 total was 3,953,593).[1] This is a surprisingly useful bit of informatio particularly for thinking about young people. How many first graders ar there? About 4 million. How many Americans under age 18? Roughly million × 18, or 72 million. Young people are about evenly divided by sex, s we can calculate that there are around 2 million 10-year-old girls, and so o
- About 2.5 million Americans die each year (there were 2,513,171 death recorded in 2011). Roughly one in four people dies of heart disease (23. percent in 2011), and cancer kills nearly as many, so that about ha (1,171,652 deaths in 2011, or 46.6 percent) die of either heart disease o cancer. In comparison, some heavily publicized causes of death are muc less common: for instance, traffic accidents killed roughly 35,000 people i 2011, breast cancer 41,000, suicide 38,000, homicide 16,000, and HIV/AID 8,000. That is, each of these causes accounted for less than 2 percent of a

deaths.[2]

- Statistics about race and ethnicity are complicated because these categorie have no precise meaning. In general, however, people who identif themselves as blacks or African Americans account for just about 13 percer of the population—about one person in eight. (Remembering that the overa population is more than 300 million, we can figure that there are about 4 million black Americans: 300 million ÷ 8 = 37.5 million.) Slightly more—ove 16 percent, or about one in six—identify themselves as Hispanic or Latin But people cannot be divided neatly into racial or ethnic categories. Mos government statistics treat Hispanic as an ethnic rather than a raci category, because Hispanics may consider themselves members of variou races. Thus, in a 2007 press release announcing that “minorities” no accounted for one-third of the U.S. population, the census burea announced that “the non-Hispanic, single-race white population [is] 6 percent of the total population.”[3] Note the awkward wording: “non Hispanic” is used because some people who classify their ethnicity a Hispanic also list their race as white; “single-race” because some peopl report mixed ancestry (such as having an American Indian ancestor). I short, the bureau is classifying as minority-group members some people wh may consider themselves white. No single, authoritative method exists fo classifying race and ethnicity. Still, a rough sense of the ethnic and racia makeup of the U.S. population can be useful.

Having this small set of basic statistical benchmarks for the overall population ca help us place the numbers we hear in context. Sometimes, when we compare statistic to these benchmarks, alarm bells may ring because a number seem improbably large or small. For instance, all other things being equal, we migh expect blacks to account for about one-eighth of people in various circumstance one-eighth of college graduates, one-eighth of prison inmates, and so on. If w learn that the actual proportion of blacks in some group is higher or lower, tha information might tell us something about the importance of race in that categor

It isn’t necessary to memorize all of these figures. They are readily availabl One of the most useful sources for basic statistics—just crammed full of offici figures—is the annual *Statistical Abstract of the United States.* It is accessibl online, and most libraries have a printed copy.[4] Whether you can remember thes basic numbers or whether you need to look them up, they can help you critical evaluate new statistics. We will have occasion to use these benchmarks (and w will identify a couple of others) later in this book.

LOOK FOR

Numbers inconsistent with benchmark figures

EXAMPLE: BATTERING DEATHS

A Web site claims that “more than four million women are battered to death by their husbands or boyfriends each year.”[5] Right away, our benchmarks help us recognize that this number can’t be correct. With only about 16,000 homicides annually, there is no chance that there could be 4 million women killed in battering incidents. In fact, 4 milli exceeds the nation’s annual 2.4 million death toll from all causes. We have no way of knowing what led the creator of the Web site to make this error, but there can be no doubt that this number is simply wrong.

Although this particular figure is clearly outlandish, I have seen it repeated on a second Web site. Statistics–both good and bad–tend to be repeated. People assume that numbers must be facts; they tell themselves that somebody must have calculated the figures, and they don’t feel obliged to check them, even against the most obvious benchmark For example, neither whoever created the 4-million-battering-deaths statistic nor the people who repeated that figure thought to ask: “Does this number for battering deaths exceed the total number of deaths from all causes?” Instead, folks feel free to repeat what they understand to be factual information. As a result, bad numbers often take on a life o their own: they continue being repeated, even after they have been thoroughly debunked. This is particularly true in th Internet age, when it is so easy to circulate information. A bad statistic is harder to kill than a vampire.

B2 Severity and Frequency

In addition to having our small set of statistical benchmarks, it is useful to keep i mind one rule of thumb: in general, the worse things are, the less common the are.

Consider child abuse and neglect. Cases of neglect far outnumber cases physical abuse, and only a small fraction of cases of physical abuse involve fat injuries. Now, one can argue that every case of child abuse and neglect is bad, bu most people would probably agree that being beaten to death is worse than, sa not having clean clothes to wear to school.

Or take crime. In 2011, there were about 700,000 motor vehicles stolen, bu fewer than 15,000 murders.[6] Stealing a car and killing someone are both bad, bu almost everyone thinks that murder is worse than car theft.

Most social problems display this pattern: there are lots of less serious case and relatively few very serious ones. This point is important because medi coverage and other claims about social problems often feature disturbin typifying examples: that is, they use dramatic cases to illustrate the problem Usually these examples are atrocity stories, chosen precisely because they ar frightening and upsetting. But this means they usually aren’t typical: mos instances of the problem are less troubling than the example. Still, it is easy t couple a terrible example to a statistic about the problem’s scope: for instance, report of an underage college student who died from acute alcohol poisoning (terrible but rare event) might be linked to an estimate of the number of underag college students who drink (doubtless a big number).[7] The implication is tha drinking on campus is a lethal problem, although, of course, the vast majority student drinkers will survive their college years.

LOOK FOR

Dramatic examples coupled to big numbers

EXAMPLE: THE INCIDENCE OF BEING INTERSEX

A person’s sex–male or female–strikes most people as the most fundamental basis for categorizing people. Classification usually occurs at the moment of birth (if not earlier, thanks to ultrasound imagery): “It’s a girl!” or “It’s a boy!” This seems so obvious and natural that most of us rarely give it a thought.

Still, there are babies who don’t fit neatly into the standard male/female framework. Some babies have ambiguous genitalia; they can be recognized as hermaphrodites at birth. Others have less visible conditions that may take years t be recognized. People with androgen insensitivity syndrome, for instance, have the XY chromosomes found in males, but because their cells do not respond to testosterone, they develop female genitalia; the condition is usually not discovered until puberty. There are several such conditions, and people with any of them may be categorized as ***intersex***.

Some advocates argue that intersex people are common enough to challenge the naturalness of the male/female distinction and that we ought to reconceptualize sex as a continuum rather than a dichotomy. Just how common is intersexuality? One widely cited estimate is that 1.7 percent of people are intersex: “For example, a city of 300,000 would have 5,100 people with varying degrees of intersexual development.”[8] (The Internet circulates claims that the actual proportion may be closer to 4 percent.)[9]

However, many of the people included in these estimates live their entire lives without discovering that they are intersex. The most common form of intersexual development is late-onset congenital adrenal hyperplasia (LOCAH–estimated to occur in 1.5 percent of all people, and therefore accounting for nearly 90 percent of all intersex individual 1.5 ÷ 1.7 = .88). Babies with LOCAH have normal genitalia that match their chromosomes; their condition may never b identified.[10] In other words, the most common variety of intersex–accounting for the great majority of cases–is subtle enough to go undiscovered. In contrast, “true hermaphrodites”–babies born with obviously ambiguous genitalia–are i fact rare; there are only about 1.2 per 100,000 births.

Intersexuality, then, displays the pattern common to so many phenomena: the most dramatic cases are relatively rare, whereas the most common cases aren’t especially dramatic.

PART 2

VARIETIES OF DUBIOUS DATA

C

BLUNDERS

Some bad statistics are the products of flubs—elementary errors. While some these mistakes might involve intentional efforts to deceive, they often refle nothing more devious than innocent errors and confusion on the part of thos presenting the numbers. For instance, after Alberta’s health minister tol students at a high school that they lived in the “suicide capital of Canada,” ministry spokesperson had to retract the claim and explain that the minister ha spoken to a doctor and “misinterpreted what they talked about.” In fact, a healt officer assured the press, the local “suicide rate is among the lowest in the regi and has been on a steady decline since the mid-1990s.”[1]

Innumeracy—the mathematical equivalent of illiteracy—affects most of us t one degree or another.[2] Oh, we may have a good grasp of the basics, such a simple arithmetic and percentages, but beyond those, things start to get a littl fuzzy, and it is easy to become confused. This confusion can affect everyone- those who produce figures, the journalists who repeat them, and the audienc that hears them. An error—say, misplacing a decimal point—may go unnoticed b the person doing the calculation. Members of the media may view their job a simply to repeat accurately what their sources say; they may tell themselves isn’t their responsibility to check their sources’ arithmetic. Those of us in th audience may assume that the media and their sources are the ones who kno about this stuff, and that what they say must be about right. And because we a have a tendency to assume that a number is a hard fact, everyone feels free t repeat the figure. Even if someone manages to correct the mistake in newspape A, the blunder takes on a life of its own and continues to crop up on TV progra B, Web site C, and blog D, which can lead still more people to repeat the error.

And yet it can be remarkably easy to spot basic blunders. In some case nothing more than a moment’s thought is enough to catch a mistake. In other our statistical benchmarks can provide a rough and ready means for checking th plausibility of numbers.

C1 The Slippery Decimal Point

The decimal point is notoriously slippery. Move it just one place to the right and- wham!—you have ten times as many of whatever you were counting. Move it jus one digit to the left and—boom!—only a tenth as many. For instance, th Associated Press reported that the final Harry Potter book sold at a magical cli on the day it was released, averaging “300,000 copies in sales per hour—mor than 50,000 a minute.”[3] Of course, the correct per-minute figure was only 5,00 copies, but this obvious mistake was overlooked not only by the reporter wh wrote the sentence but also by the editors at AP and at the various papers tha ran the story unchanged.

Misplacing a decimal point is an easy mistake to make. Sometimes our sense the world—our set of mental benchmarks—leads us to suspect that some numbe is improbably large (or small), but errors can be harder to spot when we don have a good sense of the correct number in the first place.

LOOK FOR

Numbers that seem surprisingly large–or surprisingly small

EXAMPLE: HOW MANY MINUTES BETWEEN TEEN SUICIDES?

“Today, a young person, age 14–26, kills herself or himself every 13 minutes in the United States.”–Headline on a flyer advertising a book

When I first read this headline, I wasn’t sure whether the statistic was accurate. Certainly, all teen suicide is tragic; whatever the frequency of these acts, it is too high. But could this really be happening every 13 minutes?

A bit of fiddling with my calculator showed me that there are 525,600 minutes in a year (365 days × 24 hours per da × 60 minutes per hour = 525,600). Divide that by 13 (the supposed number of minutes between young people’s suicide and we get 40,430 suicides per year. That sure seems like a lot–in fact, you may remember from our discussion of statistical benchmarks that the annual *total* number for suicides by people of all ages is only about 38,000. So right away we know something’s wrong.

In fact, government statistics tell us that there were only 4,010 suicides by young people age 15–24 in 2002 (the yea the headline appeared).[4] That works out to one every 131–not 13–minutes. Somebody must have dropped a decimal point during their calculations and, instead of producing a factoid, created what we might call a *fictoid*–a colorful but completely erroneous statistic. (Sharp-eyed readers may have noticed that, in the process, the age category 15–24 [fairly standard in government statistical reports] morphed into 14–26.)

You’ve probably seen other social problems described as occurring “every X minutes.” This is not a particularly useful way of thinking. In the first place, most of us have trouble translating these figures into useful totals, because w don’t have a good sense of how many minutes there are in a year. Knowing that there are roughly half a million– 525,600–minutes in a year is potentially useful–a good number to add to our list of benchmarks. Thus, you might say t yourself, “Hmm. Every 13 minutes would be roughly half a million divided by 13, say, around 40,000. That seems like a awful lot of suicides by young people.”

Moreover, we should not compare minutes-between-events figures from one year to the next. For instance, below the headline quoted above, the flyer continued: “Thirty years ago the suicide rate in the same group was every 26 minutes. *Why the epidemic increase?”* The problem here is that the population rises each year, but the number of minutes per year doesn’t change. Even if young people continue to commit suicide at the same rate (about 9.9 suicide per 100,000 young people in 2002), as the number of young people increases, their number of suicides will also rise, a the number of minutes between those suicides will fall. While we intuitively assume that a declining number of minute between events must mean that the problem is getting worse, that decline might simply reflect the growing populatio The actual rate at which the problem is occurring might be unchanged–or even declining.

C2 Botched Translations

It is not uncommon for people to repeat a statistic they don't actually understan Then, when they try to explain what this number means, they get it wrong, so th their innumeracy suddenly becomes visible. Or, at least it would be apparent someone understood the blunder and pointed it out.

LOOK FOR

Explanations that convert statistics into simpler language with surprising implications

EXAMPLE: MANGLING THE THREAT OF SECONDHAND SMOKE

In a press release, the British Heart Foundation's director for Scotland was quoted as saying: "We know that regular exposure to second-hand smoke increases the chances of developing heart disease by around 25%. This means that, for every four non-smokers who work in a smoky environment like a pub, one of them will suffer disability and premature death from a heart condition because of second-hand smoke."[5]

Well, no, that isn't what it means–not at all. People often make this blunder when they equate a percentage increas (such as a 25 percent increase in risk of heart disease) with an actual percentage (25 percent will get heart disease). W can make this clear with a simple example (the numbers that I am about to use are made up). Suppose that, for every 100 nonsmokers, 4 have heart disease; that means the risk of having heart disease is 4 per 100. Now let's say that exposure to secondhand smoke increases a nonsmoker's risk of heart disease by 25 percent. What's 25 percent of 4? One. So, among nonsmokers exposed to secondhand smoke, the risk of heart disease is 5 per 100 (that is, the initial ri of 4 plus an additional 1 [25 percent of 4]). The official quoted in the press release misunderstands what it means to speak of an increased risk and thinks that the risk of disease for nonsmokers exposed to secondhand smoke is 25 per 100. To use more formal language, the official is conflating relative and absolute risk.

The error was repeated in a newspaper story that quoted the press release. It is worth noting that at no point did th reporter quoting this official note the mistake (nor did an editor at the paper catch the error).[6] Perhaps they understoo that the official had mangled the statistic but decided that the quote was accurate. Or–one suspects this may be more likely–perhaps they didn't notice that anything was wrong. We can't count on the media to spot and correct every erroneous number.

Translating statistics into more easily understood terms can help us get a feel for what numbers mean, but it may also reveal that those doing the translation don't understand what they're saying.

C3 Misleading Graphs

The computer revolution has made it vastly easier for journalists not just to creat graphs but to produce jazzy, eye-catching displays of data. Sometimes the result are informative (think about the weather maps—pioneered by *USA Today*—tha show different-colored bands of temperature and give a wonderfully clear sens of the nation's weather pattern).

But a snazzy graph is not necessarily a good graph. A graph is no better tha the thinking that went into its design. And even the most familiar blunders—th errors that every guidebook on graph design warns against—are committed b people who really ought to know better.[7]

LOOK FOR

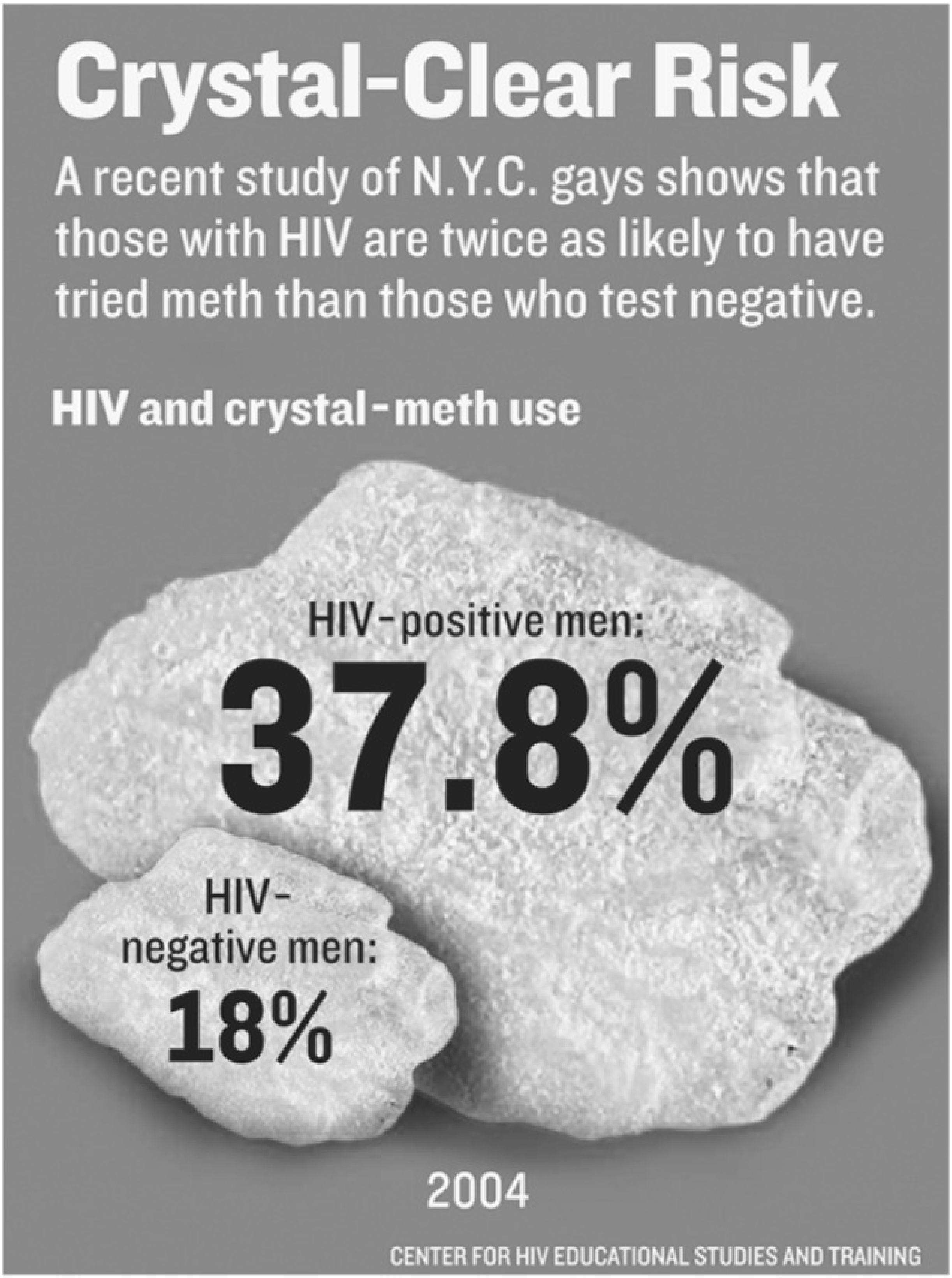
Graphs that are hard to decipher

Graphs in which the image doesn't seem to fit the data

EXAMPLE: SIZING UP METH CRYSTALS

The ~~graph shown here appeared in a major news magazine.[8] It depicts the results of a study of gay men in New~~ York City that divided them into two groups: those who tested positive for HIV, and those who tested negative. The men we asked whether they had ever tried crystal meth. About 38 percent of the HIV-positive men said they had, roughly twice the percentage (18 percent) among HIV-negative men.

Although explaining these findings takes a lot less than a thousand words, *Newsweek* decided to present them graphically. The graph illustrates findings for each group using blobs–presumably representing meth crystals. But a glance tells us that the blob/crystal for the HIV-positive group is too large; it should be about twice the size of the HIV-negative group's crystal, but it seems much larger than that.



sample content of Stat-Spotting: A Field Guide to Identifying Dubious Data

- **read Pronto**
- click Religions (Eyewitness Companions) for free
- download online Rick Steves' London 2014 pdf
- A Course in Computational Algebraic Number Theory (Graduate Texts in Mathematics, Volume 138) pdf, azw (kindle), epub
- **read online American Apostles: When Evangelicals Entered the World of Islam here**

- http://korplast.gr/lib/Pronto.pdf
- http://cambridgebrass.com/?freebooks/Religions--Eyewitness-Companions-.pdf
- http://betsy.wesleychapelcomputerrepair.com/library/Rick-Steves--London-2014.pdf
- http://ramazotti.ru/library/A-Course-in-Computational-Algebraic-Number-Theory--Graduate-Texts-in-Mathematics--Volume-138-.pdf
- http://nexson.arzamaszev.com/library/Zero-Sum--Zero-Sight--Book-2-.pdf